

Kent Academic Repository

Full text document (pdf)

Citation for published version

Li, Xinli and Li, Yijiao and Lu, Gang and Yan, Yong (2018) Biomass Fuel Identification Based on Flame Spectroscopy and Feature Engineering. Proceeding of the CSEE, 38 (15). pp. 4474-4481.

DOI

<https://doi.org/10.13334/j.0258-8013.pcsee.171984>

Link to record in KAR

<https://kar.kent.ac.uk/69883/>

Document Version

Author's Accepted Manuscript

Copyright & reuse

Content in the Kent Academic Repository is made available for research purposes. Unless otherwise stated all content is protected by copyright and in the absence of an open licence (eg Creative Commons), permissions for further reuse of content should be sought from the publisher, author or other copyright holder.

Versions of research

The version in the Kent Academic Repository may differ from the final published version.

Users are advised to check <http://kar.kent.ac.uk> for the status of the paper. **Users should always cite the published version of record.**

Enquiries

For any further enquiries regarding the licence status of this document, please contact:

researchsupport@kent.ac.uk

If you believe this document infringes copyright then please contact the KAR admin team with the take-down information provided at <http://kar.kent.ac.uk/contact.html>

DOI: 文章编号: 中图分类号: TM

基于火焰光谱和特征工程的生物质燃料识别

李新利¹, 李一娇¹, 卢钢², 闫勇^{1,2}

(1. 华北电力大学控制与计算机工程学院, 北京市 昌平区, 102206

2. 英国肯特大学工程学院, 肯特 CT2 7NT, 英国)

Biomass Fuel Identification Based on Flame Spectroscopy and Feature Engineering

LI Xin-li¹, LI Yi-jiao¹, LU Gang², YAN Yong^{1,2}

(1. School of Control and Computer Engineering, North China Electric Power University, Changping District, Beijing 102206, China;

2. School of Engineering and Digital Arts, University of Kent, Kent CT2 7NT, UK)

ABSTRACT: Flame spectra contain useful information about combustion and hence the spectral features of flame radicals may be used to identify different biomass fuels. A technique for biomass fuel identification is proposed based on the spectral features of flame radicals, feature engineering and improved support vector machine. The spectral intensity signals of biomass flames and flame radicals (OH^* (310.85nm), CN^* (390.00nm), CH^* (430.57nm) and C_2^* (515.23nm, 545.59nm)) were acquired using a spectrometer. Feature engineering was built, which can accurately reflect the characteristics of sample category, through feature extraction, feature selection based on Filter and feature learning based on dictionary learning. The support vector machine is used to build the identification model, where radial basis kernel parameter γ and error penalty factor C are optimized using an improved grid search algorithm. Experimental results from a laboratory-scale combustion rig show the effectiveness of the proposed method for the identification of biomass fuel.

KEY WORDS: flame spectroscopy; flame radicals; biomass; fuel identification; feature engineering; dictionary learning; support vector machine

摘要: 火焰光谱包含了丰富的燃烧信息, 火焰自由基的光谱特征对不同生物质燃料识别具有重要影响。文中通过生物质燃烧火焰和火焰自由基光谱特征的测量, 结合特征工程, 提出一种基于改进支持向量机的生物质燃料识别技术。该技术通过光纤光谱仪获得生物质火焰辐射强度和火焰自由基(OH^* (310.85nm), CN^* (390.00nm), CH^* (430.57nm)和 C_2^* (515.23nm、545.59nm))辐射强度信号, 通过特征

提取、基于 Filter 的特征选择和基于字典学习的特征学习, 构建特征工程, 获得能够准确反应样本类别的特征, 并结合改进的网格搜索算法优化支持向量机的径向基核参数 γ 和误差惩罚因子 C , 建立生物质燃料识别模型。在燃烧试验炉上的实验结果验证了该模型的有效性。

关键词: 火焰光谱; 火焰自由基; 生物质; 燃料识别; 特征工程; 字典学习; 支持向量机

0 引言

在燃煤或生物质锅炉燃烧中, 燃料品种多样, 实际所用燃料经常与所设计的燃料不同。燃料种类的变化会造成结焦和积灰严重等问题, 直接影响生产的安全性和经济效益^[1]。因此, 研究燃料种类的识别对锅炉安全、经济、有效燃烧有着重要意义。目前已有研究团队对燃料识别进行了研究, 但燃料识别的对象多为燃煤, 关于生物质的燃料识别研究较为有限。

传统的燃料识别方法, 通常检测其水分、灰分、发热量、挥发份等来分析燃料品质, 通过灼烧、化学、辐射等方法能够获得准确的燃料种类信息, 但这些方法多为离线, 技术繁琐、耗时且所用仪器价格昂贵^[2-3]。近年来, 随着火焰在线检测技术的发展, 国内外将目光投向基于火焰检测技术的在线燃料识别, 特别是基于图像和辐射光能的火焰检测技术^[4-8]。基于图像的火焰检测技术, 虽然其全面清晰的图像比单纯的火焰光谱包含更全面的燃烧信息, 但

其对检测镜头的清洁度以及冷却要求较高,现场应用受到一定限制。基于辐射光能的火焰检测方法利用辐射光能的原理检测炉膛的火焰,数据处理量较小,设备简单易用,在工业现场应用较广泛。文[9]采用四个光电二极管采集了不同波段(如400-1100nm, 900-1700nm, 1800-3200nm, 2000-7000nm)的火焰信号,分别提取了直流分量、交流分量、均方值、偏度、峭度和闪烁频率的时域频域特征,基于模糊算法实现生物质、单煤、混煤的识别。文[10]采用三个光电二极管构成的火焰探测器获得0.4 μ m到3.2 μ m的火焰信号,提取了包括平均值、标准偏差和闪烁频率的三个火焰信息,构成火焰特征矩阵,通过计算矩阵的相关系数识别煤种。Huang Y等^[11]提取了峰峰值、闪烁频率和闪烁强度作为火焰信号的特征量,结合最小二乘支持向量机(Least Squares Support Vector Machine, LSSVM)来识别燃料。徐立军等^[12-15]使用包含三个光电二极管的火焰探测器获得可见光、红外线和紫外线三个波段的火焰信号,分别提取了三个波段的稳态与动态分量、锐度、峭度和火焰闪烁频率等特征,采用主成分分析降维,结合神经网络、支持向量机(Support Vector Machine, SVM)、联合概率密度判别器(Joint Probability Density Arbiter, JPDA)等软计算模型来识别燃料。

在上述基于辐射光能的燃料识别研究中,基本上只采用了火焰的特定辐射波段的信号,如紫外、可见光等。燃料燃烧火焰含有丰富信息,其中火焰自由基(如OH^{*}、CN^{*}、CH^{*}和C₂^{*}),作为燃烧中间产物,不仅反映了燃烧机理,更包含了燃料品种、燃烧状况、污染物生成等大量信息。文[16-17]基于火焰自由基成像技术,提取火焰自由基(OH^{*}、CN^{*}、CH^{*}和C₂^{*})的特征,结合轮廓转换、最小二乘支持向量回归及径向基函数建模预测生物质燃烧过程中NO_x的排放。文[18-19]基于图像处理技术,通过提取不同火焰自由基的图像特征实现燃料识别。而基于辐射光能的燃料识别鲜有将燃烧中间产物——火焰自由基的信息纳入燃料识别的特征。

目前的燃料识别模型,主要从特征构建和分类算法展开研究,侧重于分类算法的改进。虽然基于历史经验和专业背景提取的特征越来越丰富,但是特征构建的内容局限在特征提取和特征降维(如主成分分析(Principal Component Analysis, PCA)、独立主元分析(Independent Component Correlation

Algorithm, ICA)。其中,特征提取是依据火焰燃烧特性的先验知识进行特征选择与指标定义,特征降维是为了降低识别模型复杂度对高维特征的一种处理手段。整体来说,在燃料识别问题上,对于反映燃料燃烧种类的特征信息还没有学者进行细致的特征工程建设。特征工程是指利用目标问题所在特定领域的知识或者自动化的方法来生成、提取、删减或者组合变化得到的特征的处理步骤,包括特征提取、选择、学习和变换等一系列子任务,能够对源数据进行充分的挖掘获得所需特征^[20,21]。其中特征提取依赖于专业领域先验知识,特征选择一般有包装(Wrapper)、嵌入(Embedded)和过滤(Filter)方法^[22]。包装法计算代价高,嵌入法将特征选择技术嵌入到模型训练中,而过滤类型方法通过一种代理评价标准评估特征,实现起来较为简单。特征学习是通过一系列机器学习方法将原始输入特征转换成新的特征表示,以便在新任务中能提高性能^[23]。通过特征工程,可以构建能够反映燃料种类的优质特征,从而建立更有效的燃料识别模型。

光纤光谱仪广泛应用于科研和工业领域,由于其测量系统的模块化和灵活性,具有检测精度高、速度快等优点,可用于火焰光谱在线测量分析。本文采用光纤光谱仪获取不同生物质燃烧的火焰光谱(200-1000nm)和燃烧中间产物火焰自由基(OH^{*}、CN^{*}、CH^{*}和C₂^{*})的光谱辐射信号,然后对光谱信号构建特征工程,通过特征提取、基于Filter的特征选择和基于字典学习的特征学习,获得能够准确反应样本类别的特征,通过改进的网格搜索算法优化支持向量机主要参数,缩短参数寻优时间,构建基于特征工程和支持向量机的生物质燃料识别模型。

1 实验技术路线

1.1 技术路线

图1为基于火焰光谱和特征工程的生物质燃料识别系统结构图。本系统通过由光纤光谱仪和计算机组成的光谱数据采集系统获取不同生物质燃料(玉米芯、柳木、花生壳和麦秸)的火焰辐射强度、火焰辐射能量、火焰自由基OH^{*}(310.85nm)、CN^{*}(390.00nm)、CH^{*}(430.57nm)和C₂^{*}(515.23nm,记为C₂^{*}-1, 545.59nm,记为C₂^{*}-2)的辐射强度信号。经过数据预处理、特征工程(包含特征提取、

基于 Filter 的特征选择和基于字典学习的特征学习) 获得反映燃料种类的特征, 建立基于支持向量机的生物质燃料识别模型。

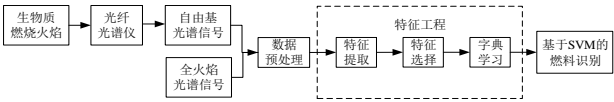


图 1 基于火焰光谱和特征工程的
生物质燃料识别系统框图

表 1 四种生物质燃料工业分析与元素分析

Tab. 1 Industrial and elemental analysis of four biomass fuels

生物质	分析结果(%)									低位发热量 (kJ/kg)
	水分	挥发分	灰分	固定碳	C	O	H	S	N	
玉米芯	5.87	74.55	1.34	18.24	39.68	59.62	<0.1	0.51	0.19	17622
柳木	9.70	69.93	3.22	17.14	44.67	49.24	5.33	0.56	0.20	18030
花生壳	7.87	70.23	4.21	17.69	43.03	49.52	5.56	0.51	1.37	17470
麦秸	10.18	58.66	17.58	13.58	32.17	63.18	3.97	0.43	0.26	15148

生物质的可燃成分为挥发分和固定碳, 水分和灰分为不可燃成分。由上表可知, 四种生物质中, 水分、灰分含量最低的是玉米芯, 最高是麦秸, 挥发分、固定碳含量最低为麦秸, 最高是玉米芯。柳木、花生壳的各个工业元素含量较为接近。从元素分析结果看出, 生物质主要由 C、H、O 及少量的 N、S 等元素组成, 其中, 柳木的 C 元素含量最高, 麦秸的 O 元素含量最高。四种生物质的柳木低位热值最高, 麦秸最低。

2 数据预处理和构建特征工程

2.1 数据预处理

数据预处理可以清理原始数据中的冗余数据, 进行数据聚集实现对杂乱数据的简单分类, 观察聚集数据直观的统计变化趋势, 进一步进行特征构建。本文通过数理统计、数据聚集(统计等间距的光谱区间内, 光谱强度的频数, 以数据集的平均值为中心, 筛选累积频率在 90% 以上的光谱数据), 去除生物质燃料燃烧过程中熄火(对应数据异常点、零点)、燃烧不稳定(对应波动剧烈数据)等火焰光谱数据, 获得四种生物质(玉米芯、柳木、花生壳和麦秸)的 4000 个样本, 每种生物质样本量为 1000 个, 每个包含火焰及火焰自由基的 7 维原始特征, 分别为 OH^{*}、CH^{*}、CN^{*}、C₂^{*}-1、C₂^{*}-2 辐射强度、火焰辐射强度和火焰辐射能量。

2.2 特征工程设计

Fig. 1 Schematic of biomass fuel identification system based on flame spectroscopy and feature engineering

1.2 生物质燃烧特性

本文选取玉米芯、柳木、花生壳、麦秸四种生物质燃料, 通过专业机构对其进行化学分析, 获得四种生物质燃料的工业分析、元素分析及发热量, 如表 1 所示。

2.2.1 特征提取

基于数理统计, 将 1000 个原始样本集中每 10 个数据作为一组样本, 提取 OH^{*}、CN^{*}、CH^{*}、C₂^{*}-1、C₂^{*}-2 辐射强度、火焰辐射强度和火焰辐射能量的时域特征, 包括标准差、方差、均值、最大值、最小值、峰峰值、峰度、偏度、波动率、比方差, 以及提取火焰辐射强度的频域特征, 包括闪烁频率和均匀度。最终获得 100 组样本, 72 维特征。

假设 x_i 为第 i 样本点的辐射强度, N 为样本点个数, X 为样本集。偏度 S 和峰度 K 分别是信号的三阶矩和四阶矩, 如式(1)、(2)所示, 反映了火焰信号概率密度函数的不对称程度和陡峭程度, 两者都是反映火焰信号中大幅度值成分的影响。

$$S = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - u}{\sigma} \right)^3 \tag{1}$$

$$K = \frac{1}{N} \sum_{i=1}^N \left(\frac{x_i - u}{\sigma} \right)^4 \tag{2}$$

其中, u 为均值, σ 为标准差。

波动率 R: 火焰闪烁时最大值偏离最小值的程度, 即,

$$R = \frac{\text{MAX}(X) - \text{MIN}(X)}{\text{MEAN}(X)} \tag{3}$$

比方差 B: 火焰强度值的方差和均值之比。

$$B = \frac{D}{u} \tag{4}$$

其中, D 为方差。

闪烁频率 F^[24]: 火焰闪烁的加权平均频率, 反

映了火焰脉动能力大小的量。其值越大,表明火焰的脉动能力越强,火焰越稳定。其中, $f(i)$ 、 $P(i)$ 分别为第 i 个点的频率大小和功率谱值。

$$F = \frac{\sum_{i=1}^N [P(i) \times f(i)]}{\sum_{i=1}^N P(i)} \quad (5)$$

均匀度 U : 频谱图中低频波能量变大时,预示着燃烧的稳定性变差。域内提取“均匀度”作为表征稳定性的一个特征量。其中, R 为功率谱中后半高频部分功率谱值之和, L 为功率谱中前半低频部分功率谱值之和^[25]。

$$U = \frac{R}{L} \quad (6)$$

表 2 OH*强度的 10 维特征相关系数矩阵

Tab. 2 Correlation coefficient matrix of 10 - dimensional features of OH* intensity

	标准差	方差	均值	最小值	最大值	峰峰值	峰度	偏度	波动率	比方差
标准差	1.00	1.00	0.54	0.38	0.72	0.97	-0.01	0.27	0.95	0.99
方差	1.00	1.00	0.54	0.38	0.72	0.97	-0.01	0.27	0.95	0.99
均值	0.54	0.54	1.00	0.93	0.93	0.55	0.04	0.01	0.35	0.44
最小值	0.38	0.38	0.93	1.00	0.87	0.38	0.03	0.18	0.17	0.28
最大值	0.72	0.72	0.93	0.87	1.00	0.73	0.08	0.24	0.56	0.64
峰峰值	0.97	0.97	0.55	0.38	0.73	1.00	0.16	0.25	0.97	0.96
峰度	-0.01	-0.01	0.04	0.03	0.08	0.16	1.00	0.11	0.17	-0.02
偏度	0.27	0.27	0.01	0.18	0.24	0.25	0.11	1.00	0.27	0.29
波动率	0.95	0.95	0.35	0.17	0.56	0.97	0.17	0.27	1.00	0.96
比方差	0.99	0.99	0.44	0.28	0.64	0.96	-0.02	0.29	0.96	1.00

采用同样的方法,分别获得 CH^* 、 CN^* 、 C_2^*-1 、 C_2^*-2 辐射强度和火焰辐射能量的包括标准差、均值、最小值、最大值、峰度、偏度和波动率 7 维特征,合计 42 维特征。对于火焰辐射强度,计算其 12 维提取特征的 Spearman 相关系数,设置阈值为 0.96,去除火焰强度的方差、峰峰值、比方差和均匀度共 4 维特征,选择标准差、均值、最小值、最大值、峰度、偏度、波动率和闪烁频率共 8 维特征。

2.2.3 基于字典学习的特征学习

近年来,压缩感知及稀疏编码受到越来越多研究者的关注,由 Wright 等提出的基于稀疏表示的分类 (Sparse representation based classification, SRC) 为模式分类问题开辟了一条崭新的途径,在图像目标识别和分类 (如人脸识别) 方面取得了显著的效果^[27,28]。在稀疏表示的基础上,字典学习方法通过对输入的数据集学习一个自适应的字典,可以得到原始数据集精确的特征表示,对字典加以鉴别性约束,使得学习的字典能更好地进行分类。

稀疏表示的字典通过训练样本学习得到,它可以是超完备 (over-completed) 的,其系数求解仅需满足稀疏性约束,这使得表示精度大大提高,从而

2.2.2 基于 Filter 的特征选择

对于特征提取的 72 维特征,本文采用相关性度量进行特征选择,去除冗余和不相关特征,形成特征子集。相关性度量采用 Spearman^[26]相关系数法,通过设定阈值筛选相关性较小的特征。

以 OH^* 辐射强度为例,计算其 10 维提取特征的 Spearman 相关系数,得到相关系数矩阵如表 2。结合特征参数的物理含义,设置阈值为 0.95,进行特征选择,去除 OH^* 强度的方差、峰峰值和比方差 3 维特征,选择标准差、均值、最小值、最大值、峰度、偏度和波动率共 7 维特征。

提升了利用表示残差进行分类的分类精度。字典学习方法分为无监督学习和有监督学习。虽然无监督的字典学习算法应用于分类中并取得一定的效果,但最近的研究表明通过有监督的字典学习能够学到更适合分类的字典^[29,30]。本文采用有监督的字典学习,以样本的类别标签作为监督进行训练。

假设有 N 类样本,训练样本矩阵可以表示为

$$X = [X_1, X_2, \dots, X_N], \quad X_i \text{ 表示 } X \text{ 的第 } i \text{ 类样本, 并}$$

且 $X_i \in \mathbb{R}^{\dim \times N_i}$, N_i 表示第 i 类样本的个数。同样字典

可以表示为 $D = [D_1, D_2, \dots, D_N]$, 表示第 i 类字典

典。训练样本 X 可以表示为字典 D 中若干字典原子的线性组合: $X \approx DG$ 。其中 G 为训练样本在字典 D 学习下的稀疏表示系数。字典学习表示如下:

$$\min_{D, G} \sum_{i=1}^N \|X - DG\|_2^2 + \lambda \sum_{i=1}^N \|G\|_1 + E_{\text{discriminant}}(D, G, Y) \quad (7)$$

其中 λ 表示正则化系数, Y 是训练样本 X 的类别矩

阵: $\|X - DG\|_2^2$ 是为了确保学习到的字典 D 对训练

样本 X 的重构能力; $\|G\|_1$ 表示 G 的 L_1 范数, 是对系

数 G 的稀疏性约束; $E_{\text{discriminant}}(D, G, Y)$ 是鉴别约束

项, 使得学习的字典 D 具有鉴别性。求解该问题一般使用交替优化的方法, 先初始化字典 D , 固定字典 D 对 G 进行优化, 然后固定 G 对字典 D 进行优化。重复上述两步, 不断交替的优化 D 和 G 直至收敛。

分类阶段采取残差最小的准则, 即分别计算各类的残差, 将测试样本划分为残差最小的那一类。给定测试样本 y , 分类如下式所示:

$$\text{identify}(y) = \arg \min_i \{e_i\} \quad (8)$$

其中, $e_i = \|y - D_i g_i\|_2$, $g_i (i = 1, 2, \dots, N)$ 为测试样

本 y 在子字典 D_i 下的稀疏表示系数, $\arg \min_i \{e_i\}$ 表

示使 e_i 达到最小值时 i 的取值。

3 基于 SVM 的生物质燃料识别模型

支持向量机的主要思想是建立一个分类超平面作为决策面^[31], 使得正例和反例之间的隔离边缘最大化, 从而实现样本分类。当数据是线性可分时, 可以计算超平面。在数据是线性不可分的情况下, 数据被映射到一个高维空间, 建立一个分离超平面。最终, 分类问题转化为求解二次规划约束优化问题。

求解的极小值函数为:

$$\begin{aligned} \Phi(w, \xi) &= \frac{1}{2}(w \cdot w) + C \left(\sum_{i=1}^n \xi_i \right) \\ \text{s.t.} \quad &\begin{cases} y_i [(w \cdot x_i) + b] \geq 1 - \xi_i, i = 1, 2, \dots, l \\ \xi_i \geq 0 \end{cases} \end{aligned} \quad (9)$$

其中, w 为超平面的法向量; b 为超平面的平移量; C 为惩罚系数, 控制系数 ξ 边界的平衡。

引入拉格朗日乘子 a_i , 根据泛函理论, 在满足

核函数 $K(x \cdot x_i) = \varphi(x) \cdot \varphi(x_i)$ 条件下, 得到分类函

数为:

$$\begin{aligned} f(x) &= \text{sgn}((w \cdot x) + b) \\ &= \text{sgn}(\sum_{i=1}^k a_i \times y_i K(x \cdot x_i) + b) \end{aligned} \quad (10)$$

其中, 本文选择为 RBF 核函数, 如式 (11) 所示。

$$K(x \cdot x_i) = \exp\left(-\frac{|x - x_i|^2}{\gamma^2}\right), \gamma > 0 \quad (11)$$

在支持向量机算法中, 径向基核参数 γ 和误差惩罚因子 C 是两个重要的参数, 其选择的好坏直接影响着支持向量机的性能。较为常用的参数寻优方法有网格搜索、遗传算法、粒子群算法等。网格搜索遍历参数构成的网格区间来寻找最优值, 时间成本非常高。遗传算法和粒子群算法为启发式算法, 虽然无需遍历区间内的参数组, 但实现较为复杂, 容易陷入局部最优。本文采用改进的网格搜索算法寻找 LIBSVM 最优参数, 先采用大步矩大范围的粗搜, 初步确定一个最优参数区间, 之后在此区间内进行小步距精搜, 不仅实现起来简单, 而且大大的减少了参数寻优的时间, 能够进行更有效的燃料识别。

4 试验结果与分析

4.1 试验装置和工况

试验装置 (如图 2 所示) 由空气压缩机、送风管路、燃气管路、电子质量流量计、浮子流量计、燃烧室及生物质给料装置组成。甲烷和空气从各自的气路管道输送至燃烧器, 在燃烧器中预先混合, 点燃后产生预混火焰。生物质 (玉米芯、柳木、花生壳和麦秸) 由进料螺杆旋转推送并通过给料空气将粉体送达燃烧反应区进行燃烧。

试验中甲烷流速为 1l/min, 空气和给料空气的流速为 5l/min, 生物质给料器进料频率为 3Hz, 进料质量流量为 1.295g/min。光纤光谱仪测量火焰光谱强度, 通过设置光谱仪参数, 去除环境光、暗噪声及光学元件缺陷等因素引起的误差, 减少仪器本身和周围环境噪声对测量精度的影响。试验中首先采集背景辐射光谱, 测量时直接扣除背景辐射, 从而减少背景辐射误差。每次测量值为 3 次光谱扫描的平均值, 且平滑度为 3 (将任意时间点的光谱数

据和其前后的 3 个数据点做平均,平均后的数值为该时间点的光谱辐射强度),从而降低噪声,提高信噪比,降低火焰跳动、杂波信号等对光谱信号检测的干扰。

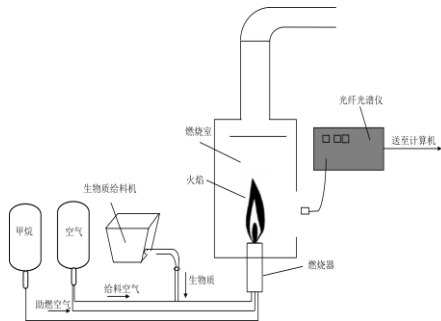


图 2 试验装置

Fig. 2 Experimental setup

4.2 生物质燃料识别模型构建

4.2.1 数据处理和特征工程

试验在稳定燃烧状况下,使生物质燃料充分燃烧,以减少碳烟等颗粒对自由基的信号采集影响。试验采集玉米芯、柳木、花生壳和麦秸四种不同生物质燃料的火焰自由基(OH^* 、 CN^* 、 CH^* 、 C_2^*-1 和 C_2^*-2)的辐射强度信号、火焰辐射强度和火焰辐射能量,经过数据预处理,每种生物质有 1000 个原始样本,每组样本包含 7 维原始特征。将原始样本集中每 10 个数据作为一组新的样本,提取各个原始特征在时域和频域的特征,获得包括火焰辐射强度的 12 维特征,火焰辐射能量、 OH^* 、 CN^* 、 CH^* 、 C_2^*-1 和 C_2^*-2 辐射强度各 10 维特征,共计 72 维特征,100 组样本。通过 Filter 的 Spearman 相关系数法,去除冗余、相关系数高的特征,获得包括火焰辐射强度的 8 维特征,火焰辐射能量、 OH^* 、 CN^* 、 CH^* 、 C_2^*-1 和 C_2^*-2 辐射强度各 7 维特征,合计共 50 维特征。经过特征提取和特征选择,最终获得 50 维特征,100 组样本。

将特征提取、特征选择获得每种生物质的 100*50 维特征,进行有监督的字典学习,获得了 100*200 维的稀疏表示特征向量,作为燃料识别模型的输入。

4.2.2 基于网格搜索的支持向量机

随机选取每种生物质 100 组样本的 70%作为训练集,即玉米芯、柳木、花生壳和麦秸四种生物质各选取 70 组样本,共 280 组样本进行训练。样本集剩余的 30%样本作为测试集,测试燃料识别效果。

试验所用的计算机配置为处理器 Intel(R) Core(TM)2 Duo CPU,内存 2.00GB,64 位操作系统。

设定网格搜索的参数变量(C, γ)的范围及步距,其中 C 的初始范围设置为 $[2^{-10}, 2^{10}]$, γ 的初始范围为 $[2^{-10}, 2^{10}]$,步距设置为 0.1,得到分别使得训练集和测试集的识别准确率均为 100%的全局最优参数为 $C=1.3195$, $\gamma=1.8661$,但是耗费的时间较长,为 4655.19 秒。由参数选择结果的等高线图(如图 3 所示)可以看到使得训练集识别准确率为 80%以上的(C, γ)的范围大致在 $([2^{-10}, 2^5], [2^{-3}, 2^{10}])$,且使得识别准确率最优的不仅仅为单一的参数组合,而是一组参数组合,可为小步距搜索提供依据。

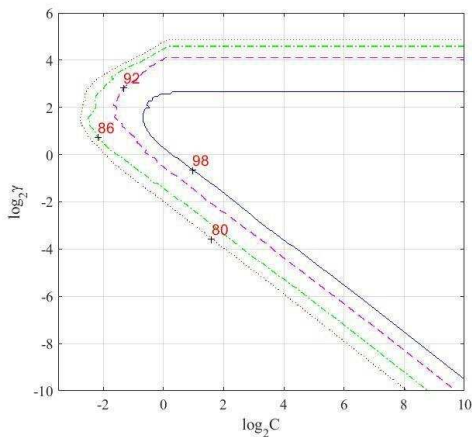


图 3 参数选择等高线图

Fig.3 Contours of parameter selection.

改进的网格搜索算法,设定网格搜索的变量(C, γ)的范围及步距,其中 C 的初始范围设置为 $[2^{-10}, 2^{10}]$, γ 的初始范围为 $[2^{-10}, 2^{10}]$,一般传统的步距一般为 0.1,改进的方法中初始步距选为 100 倍的步距,设为 10。

采用 K 折交叉验证(K-fold Cross Validation, K-CV)方法对训练集进行测试,其中 $K=3$,得到使识别准确率最高的局部最优参数 $C=2^{10}$ 和 $\gamma=1$,耗费时间为 1.45 秒。

根据得到的局部最优参数,在其附近选择不同的区间进行二次寻优,步距与传统方法一样设为 0.1。比较不同的搜索区间对测试集识别准确率的影响,如表 3 所示。

表 3 不同搜索区间内参数寻优的结果比较

Tab. 3 Comparison of parameters in different search

intervals			
参数	第 1 组	第 2 组	第 3 组
C 范围	2^9-2^{10}	2^8-2^{10}	2^7-2^{10}
γ 范围	$2^{-1}-2^1$	$2^{-2}-2^2$	$2^{-3}-2^3$

Best C	512	256	128
Best γ	0.5	0.25	0.125
准确率(%)	100	100	100
时间(s)	15.14	63.13	145.68

由上表可知,当参数的范围为(2^9 - 2^{10} , 2^{-1} - 2^{-1})时,可以获得 100%的准确识别率,且时间成本最低,加上大步距范围粗搜的时间成本 1.45 秒,所需时间为 16.59 秒,远低于未改进的网格搜索算法的耗时(4655.19 秒),在保持高识别率的前提下,降低了参数寻优的时间成本。

5 结论

采用光纤光谱仪获得不同生物质燃料的火焰及火焰自由基 OH^* , CN^* , CH^* 和 C_2^* 的光谱辐射强度信号,构建特征工程,对基础数据进行深入挖掘,建立了基于特征提取、Filter 的特征选择和字典学习的特征学习方法以获得反映生物质燃料种类的优质特征。采用改进的网格搜索算法优化支持向量机的径向基核参数 γ 和误差惩罚因子 C ,构建了基于 LIBSVM 的生物质燃料识别模型,在燃烧试验炉上的实验结果验证了该模型的有效性,且时间成本远远低于改进前的支持向量机模型,是一种有效的生物质燃料识别模型。

参考文献

- [1] 谭厚章, 苗杨, 王洋, 王学斌, 徐通模. 煤质多变下火电厂最经济煤种决策[J]. 中国电机工程学报, 2009, 29(14):1-5
Tan Houzhang, Miao Yang, Wang Yang, Wang Xuebin, Xu Tongmo. Decision of the Most Economic Coal in Power Plant With Coal Quality Variation[J]. Proceedings of the CSEE, 2009, 29(14):1-5
- [2] 梁勇. 论近红外光谱技术在煤质检测分析中的应用[J]. 辽宁化工, 2017, 46(3): 312-314.
Liang Yong. Application of Near Infrared Spectroscopy in Coal Quality Detection and Analysis[J]. Liaoning Chemical Industry, 2017, 46(3): 312-314.
- [3] 赵忠辉. 煤质在线检测和机械化采制样一体化技术研究[J]. 煤炭技术, 2016, 35(4): 236-238.
Zhao Zhonghui. Research on Integration Technology of Coal Quality On-line Detection and Mechanical Sampling[J]. Coal Technology, 2016, 35(4): 236-238.
- [4] Min G K, Sung B P. A study on the flame detection and object classification technique using the color information[C]. 2015 10th International Conference for Internet Technology and Secured Transactions (ICITST), London, UK, 2015, 120-123.
- [5] Chen H, Zhang X G, Hong P Y, Hu H P, Yin X. Recognition of the Temperature Condition of a Rotary Kiln Using Dynamic Features of a Series of Blurry Flame Images[J]. IEEE Transactions on Industrial Informatics, 2016, 12(1): 148-157
- [6] Liu J X, Yang B, Cai X S. Research of combustion Diagnostics and testing in boiler based on radiation spectroscopy[J]. Journal of Engineering Thermophysics, 2015, 36(9): 2055-2059.
- [7] Khodabakhsh A, Qu Z C, Alrahman C A, Johansson A C, Rutkowski L, Schmidt F M, Foltynowicz A. Cavity-enhanced optical frequency comb spectroscopy of high-temperature water in a flame[C]. 2015 Conference on Lasers and Electro Optics (CLEO), San Jose, CA, USA, 2015: 1-2.
- [8] Garcés H O, Rojas A J, Valdebenito V, Navarro A, Pereira C. Radiation sensor in a oil boiler based on flame spectral analysis[C]. 2016 IEEE SENSORS, Orlando, FL, USA, 2016: 1-3.
- [9] Chi T Y, Zhang J H, Yan Y, Zhou H L, Zheng H. Investigations into the ignition behaviors of pulverized coals and coal blends in a drop tube furnace using flame monitoring techniques[J]. Fuel, 2010, 89(3): 743-751.
- [10] Kan R, Zhang H J, Zhou H L, Chi T Y. A Fast Method of On-line Coal Identification in a Drop-tube Furnace[C]. Instrumentation and Measurement Technology Conference, Singapore, Singapore, 2009: 234-239.
- [11] Huang Y S, Liu S, Li J, et al. Fuel Identification Based on the Least Squares Support Vector Machines[J]. Advanced Materials Research, 2011, 317-319: 1237-1240.
- [12] Xu L J, Yan Y, Cornwel S, et al. Online Tracking by Combing Principal Component Analysis and Neural Network Techniques[J]. IEEE Transactions on Instrumentation and Measurement, 2005, 54(4): 1640-1645.
- [13] Tan C, Xu L J, Cao Z. On-line fuel identification using optical sensing and support vector machines technique[C]. IEEE Instrumentation and Measurement Technology Conference, Singapore, Singapore, 2009: 1144 - 1147.
- [14] Xu L J, Tan C, Li X M, et al. Fuel-type Identification using Joint Probability Density Arbiter and Soft-Computing Techniques[J]. IEEE Transactions on Instrumentation and Measurement, 2012, 61(2): 286-296.
- [15] Tan C, Xu L J, Li X M, et al. Independent Component Analysis-based Fuel Type Identification for Coal-fired Power Plant[J]. Combustion Science and Technology,

- 2012, 184(3): 277-292.
- [16] Li N, Lu G, Li X L, Yan Y. Prediction of Pollutant Emissions of Biomass Flames Through Digital Imaging, Contourlet Transform, and Support Vector Regression Modeling[J]. IEEE Transactions on Instrumentation and Measurement, 2015, 64(9): 2409-2416.
- [17] Li N, Lu G, Li X L, Yan Y. Prediction of pollutant emissions of biomass flames using digital imaging, contourlet transform and Radial Basis Function network techniques[C]. Instrumentation and Measurement Technology Conference (I2MTC) Proceedings, Montevideo, Uruguay, 2014: 697-700.
- [18] Li X L, Wu M J, Lu G, Yan Y, Liu S. On-line identification of biomass fuels based on flame radical imaging and application of radical basis function neural network techniques[J]. IET Renewable Power Generation, 2015, 9(4): 323-330.
- [19] Li X L, Li N, Lu G, Yan Y. On-line identification of biomass fuels based on flame radical and application of Support Vector Machine techniques[J]. 2nd IET Renewable Power Generation Conference, Beijing, China, 2013: 1-4.
- [20] Anderson M, Antenucci D, Bittorf V, et al. Brainwash: a data system for feature engineering[C]. Proc of Biennial Conference on Innovative Data Systems Research. 2013: 100-104.
- [21] Yang Shulin. Feature engineering in fine-grained image classification[D]. Washington DC: University of Washington, 2013.
- [22] Guyon I, Elisseeff A. An introduction to variable and feature selection[J]. Journal of Machine Learning Research, 2003, 3(1): 1157-1182.
- [23] Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives[J]. IEEE Trans on Pattern Analysis and Machine Intelligence, 2013, 35(8): 1798-1828.
- [24] Huang Y, Yan Y, Lu G, Reed A. On-line flicker measurement of gaseous flames by image processing and spectral analysis[J]. Meas Sci Technol, 1999, 10(8): 726.
- [25] Prathap C, Ray Anjan, Ravi M R. Investigation of nitrogen dilution effects on the laminar burning velocity and flame stability of syngas fuel at atmospheric condition[J], Combustion and Flame, 2008, 155(1-2): 145-160.
- [26] Zhang W Y, Wei Z W, Wang B H, Han X P. Measuring mixing patterns in complex networks by Spearman rank correlation coefficient[J]. Physica A: Statistical Mechanics and its Applications, 2016, 451: 440-450.
- [27] Wright J, Yang A Y, Ganesh A, Sastry S S. Robust face recognition via sparse representation[J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2009, 31(2): 210-227.
- [28] 马小虎, 谭延琪. 基于鉴别稀疏保持嵌入的人脸识别算法[J]. 自动化学报, 2014, 40(1): 73-82.
- Ma Xiao-Hu, Tan Yan-Qi. Face recognition based on discriminant sparsity preserving embedding. Acta Automatica Sinica, 2014, 40(1): 73-82.
- [29] Kavukcuoglu K, Ranzato M, Fergus R, et al. Learning Invariant Features through Topographic Fliter Maps[J]. IEEE Conference on Computer Vision and Pattern Recognition, 2009: 1605-1612.
- [30] Yang M, Zhang L, Feng X, et al. Fisher Discrimination Dictionary Learning for Sparse Representation[J]. IEEE International Conference on Computer Vision, 2011: 543-550.
- [31] Cortes C, Vapnik V. Support-vector network[J]. Machine Learning, 1995, 20(3): 273-297.



作者简介:

李新利(1973),女,工学博士,副教授,主要从事模式识别与智能系统、图像处理等方向研究工作,

lixinli@ncepu.edu.cn

